Making Almost Commuting Matrices Commute

M. B. Hastings
Microsoft Research, Station Q, Elings Hall, University of California, Santa Barbara, CA 93106

Suppose two Hermitian matrices A, B almost commute ($\|[A, B]\| \leq \delta$). Are they close to a commuting pair of Hermitian matrices, A', B', with $\|A - A'\|, \|B - B'\| \leq \epsilon$? A theorem of H. Lin[3] shows that this is uniformly true, in that for every $\epsilon > 0$ there exists a $\delta > 0$, independent of the size N of the matrices, for which almost commuting implies being close to a commuting pair. However, this theorem does not specify how δ depends on ϵ. We give uniform bounds relating δ and ϵ. The proof is constructive, giving an explicit algorithm to construct A' and B'. We provide tighter bounds in the case of block tridiagonal and tridiagonal matrices. Within the context of quantum measurement, this implies an algorithm to construct a basis in which we can make a *projective* measurement that approximately measures two approximately commuting operators simultaneously. Finally, we comment briefly on the case of approximately measuring three or more approximately commuting operators using POVMs (positive operator-valued measures) instead of projective measurements.

The problem of when two almost commuting matrices are close to matrices which exactly commute, or, equivalently, when a matrix which is close to normal is close to a normal matrix, has a long history. See, for example [1, 2], and other references in [3] where it is mentioned that the problem dates back to the 1950s or earlier. Finally in 1995, Lin[3] proved that for any $\epsilon > 0$, there is a $\delta > 0$ such that for all N, for any pair of Hermitian N-by-N matrices, A, B, with $\|A\|, \|B\| \leq 1$, and $\|[A, B]\| \leq \delta$, there exists a pair A', B' with $[A', B'] = 0$ and $\|A - A'\| \leq \epsilon$ and $\|B - B'\| \leq \epsilon$ This proof was later shortened and generalized by Friis and Rordam[4]. Interestingly, the same is not true for almost commuting unitary matrices[5] or for almost commuting triplets[6, 7].

The importance of the above results is that the bound is uniform in N. That is, δ depends only on ϵ. Unfortunately, the proofs do not give any bounds on how δ depends on ϵ. Further, the proofs of Lin and Friis and Rordam are nonconstructive, so there is no known way to find the matrices A' and B'. In this paper, we present a construction of matrices A' and B' which enables us to give lower bounds on how small δ must be to obtain a given error ϵ.

Specifically, we prove that

Theorem 1. *Let A and B be Hermitian, N-by-N matrices, with $\|A\|, \|B\| \leq 1$. Suppose $\|[A, B]\| \leq \delta$. Then, there exist Hermitian, N-by-N matrices A' and B' such that*

1: $[A', B'] = 0$.

2: $\|A' - A\| \leq \epsilon(\delta)$ *and* $\|B' - B\| \leq \epsilon(\delta)$, *with*

$$\epsilon(\delta) = E(1/\delta)\delta^{1/5}, \tag{1}$$

where the function $E(x)$ grows slower than any power of x. The function $E(x)$ does not depend on N.

Throughout this paper, we use $\|...\|$ to denote the operator norm of a matrix, and $|....|$ to denote the l^2-norm of a vector.

The proof of theorem (1) involves first constructing a related problem involving a block tridiagonal matrix, H, and a block identity matrix X (we use the term "block identity matrix" to refer to a block diagonal matrix that is proportional to the identity matrix in each block). For such matrices we prove the theorem

Theorem 2. *Let X be a block identity Hermitian matrix and let H be a block tridiagonal Hermitian matrix, with the j-th block of X equal to $c + j\Delta$ times the identity matrix, for some constants c and Δ. Let $\|H\|, \|X\| \leq 1$. Then, there exist Hermitian matrices A' and B' such that*

1: $[A', B'] = 0$.

2: $\|A' - H\| \leq \epsilon'(\Delta)$ *and* $\|B' - X\| \leq \epsilon'(\Delta)$, *with*

$$\epsilon'(\Delta) = E'(1/\Delta)\Delta^{1/4}, \tag{2}$$

where the function $E'(x)$ grows slower than any power of x. The function $E'(x)$ does not depend on the dimension of the matrices.

After proving these results, we prove a tighter bound in the case where H is a tridiagonal matrix, rather than a block tridiagonal matrix:

Theorem 3. *Let X be a diagonal Hermitian matrix and let H be a tridiagonal Hermitian matrix, with the j-th diagonal entry of X equal to $c + j\Delta$, for some constants c and Δ. Let $\|H\|, \|X\| \leq 1$. Then, there exist Hermitian matrices A' and B' such that*

1: $[A', B'] = 0$.

2: $\|A' - H\| \leq \epsilon'(\Delta)$ *and* $\|B' - X\| \leq \epsilon'(\Delta)$, *with*

$$\epsilon'(\Delta) = E''(1/\Delta)\Delta^{1/2}, \tag{3}$$

where the function $E''(x)$ grows slower than any power of x. The function $E''(x)$ does not depend on the dimension of the matrices.

The proofs rely heavily on ideas relating to Lieb-Robinson bounds[8–11]. These bounds, combined with appropriately chosen filter functions, have been used in recent years in Hamiltonian complexity to study the dynamics and ground states of quantum systems, obtaining results such as a higher dimensional Lieb-Schultz-Mattis theorem[9], a proof of exponential decay of correlations[14], studies of dynamics out of equilibrium[15–17], new algorithms for simulation of quantum systems[18–22], an area law for entanglement entropy for general interacting systems[23], study of harmonic lattice systems[24], a Goldstone theorem with fewer assumption[25], and many others. The present paper represents a different application, to the study of almost commuting matrices.

Before beginning the proof, we give some discussion of physics intuition behind the result. The next few paragraphs are purely to motivate the problems from a physics viewpoint. In the last section on quantum measurement and in the discussion at the end we give additional applications to quantum measurement and construction of Wannier functions. The section on quantum measurement is intended to be self-contained. As mentioned, we begin by relating this problem to the study of block tridiagonal matrices. We then interpret the matrix H as a Hamiltonian for a single particle moving in one dimension, and apply the Lieb-Robinson bounds. The result (2) implies that we can construct a complete orthonormal basis of states which are simultaneously localized in both position (X) and energy (H). It is certainly easy to construct an overcomplete basis of states which is localized in both position and energy, by considering, for example, Gaussian wavepackets. The interesting result is the ability to construct an orthonormal basis which satisfies this.

Additional physics intuition can be obtained by considering the case where H is a tridiagonal matrix with 0 on the diagonal and elements just above and below the diagonal equal to 1, and where X is a diagonal matrix with entries $1/N, 2/N, \ldots$. We refer to this as a uniform chain. In the uniform chain case, if we define a new matrix H' by randomly perturbing H, replacing each diagonal element of H with a small diagonal number chosen at random, the eigenvectors of H' are localized with high probability[26, 27]. Then, we can construct a matrix X' which exactly commutes with H' as follows: if v is an eigenvector of H', we choose it to have eigenvalue for X' equal to (v, Xv). Then, since the eigenvectors are localized, we find that $\|X - X'\|$ is small. The difference $\|X - X'\|$ depends on the localization length which depends inversely on the amount of disorder, while the difference $\|H - H'\|$ depends on the amount of disorder. Unfortunately, we do not have a good enough understanding of the effect of disorder for matrices H which are block tridiagonal, rather than just tridiagonal, to turn this approach into a proof for general H and X, and thus we rely on an alternative, constructive approach.

I. PROOF OF MAIN THEOREM

We now outline the proof of theorem (1). The proof is constructive, and is described by the following algorithm:

1: Construct H from A as described in section (II A) and lemma (1). We will bound $\|H - A\|$.

2: Construct X from B as described in section (II B). We will bound $\|X - B\|$. In a basis of eigenvalues of X, the matrix H will be block tridiagonal.

3: Construct a new basis as described in section (III) such that in this basis H is close to a block diagonal matrix. That is, we will bound the operator norm of the block off-diagonal part of H. The blocks will be different from the blocks considered in step **(2)** above and will be larger. Further, we will show that X is close to a block identity matrix in this basis.

4: Set A' to be the block diagonal part of H in the basis constructed in step **(3)** and set B' to the block identity matrix constructed in step **(3)**, so that $[A', B'] = 0$.

This algorithm involves several choices of constants. In a final section, (V), we indicate how to pick the constants to obtain the error bound (1). The key step will be step 3.

II. REDUCTION TO BLOCK TRIDIAGONAL PROBLEM

The first two steps of the proof above **(1,2)** reduce theorem (1) to theorem (2), while the last two steps **(3,4)** prove theorem (2). In this section we present the first two steps.

A. Construction of Finite-Range H

We begin by constructing matrix H as given in the following lemma, where the constant Δ will be chosen later.

Lemma 1. *Given Hermitian matrices A and B, with $\|[A,B]\| \leq \delta$, for any Δ there exists a Hermitian matrix H with the following properties.*

1: $\|[H,B]\| \leq \delta$.

2: *For any two vectors v_1, v_2 which are eigenvectors of B with corresponding eigenvalues x_1, x_2, and with $|x_1 - x_2| \geq \Delta$, we have $(v_1, Hv_2) = 0$.*

3: $\|A - H\| \leq \epsilon_1$, *with $\epsilon_1 = c_0\delta/\Delta$, where c_0 is a numeric constant given below.*

Proof. We define

$$H = \Delta \int \mathrm{d}t \exp(iBt) A \exp(-iBt) f(\Delta t), \tag{4}$$

where the function $f(t)$ is defined to have the Fourier transform

$$\begin{aligned} \tilde{f}(\omega) = (1-\omega^2)^3, \quad |\omega| \leq 1 \\ \tilde{f}(\omega) = 0, \quad |\omega| \geq 1, \end{aligned} \tag{5}$$

and hence the Fourier transform of $f(\Delta t)$ is supported on the interval $[-\Delta, \Delta]$. Properties (1) and (2) follow immediately from Eq. (4). Property (3) follows from

$$\begin{aligned} \|H - A\| &= \|\Big(\Delta \int \mathrm{d}t\, \exp(iBt) A \exp(-iBt) f(\Delta t)\Big) - A\| \\ &= \|\Delta \int \mathrm{d}t\, \Big(\exp(iBt) A \exp(-iBt) - A\Big) f(\Delta t)\| \\ &\leq \Delta \int \mathrm{d}t\, \|\Big(\exp(iBt) A \exp(-iBt) - A\Big)\|\; |f(\Delta t)| \\ &\leq \Delta \int \mathrm{d}t\, |t| \|[A,B]\|\; |f(\Delta t)| \\ &\leq \delta\Delta \int \mathrm{d}t\, |t f(\Delta t)| \\ &= c_0\delta/\Delta, \end{aligned} \tag{6}$$

where we define the constant c_0 by

$$c_0 = \int \mathrm{d}t |t f(t)|. \tag{7}$$

The second line in Eq.(6) follows because $\tilde{f}(0) = 1$ so that $\Delta \int \mathrm{d}t A f(\Delta t) = A$. Note that since the first and second derivatives of $\tilde{f}(\omega)$ vanish at $\omega = \pm 1$, the function $f(t)$ decays as $1/t^3$ for large t and hence c_0 is finite. Since $\tilde{f}$ is an even function, H is Hermitian.

Note that the precise form of the function $f(t)$ is unimportant: all we require is that $\tilde{f}(0) = 1$; that $\tilde{f}$ is supported on the interval $[-1,1]$; that $\tilde{f}$ is sufficiently smooth that $f(t)$ decays fast enough for the integral over t (7) to converge; and that $\tilde{f}$ is an even function. $\square$

Remark: In a basis of eigenvectors of B, property (3) in the above lemma implies that H is "finite-range", in that the off-diagonal elements are vanishing for sufficiently large $|x_1 - x_2|$. The next theorem is a Lieb-Robinson bound for such finite range Hamiltonians, similar to those proven for many-body Hamiltonians[8–11]. This result is also similar to results on the decay of entries of smooth functions of matrices proven in [12, 13].

We now introduce some terminology. Given two sets of real numbers, S_1, S_2, we define

$$\mathrm{dist}(S_1, S_2) = \min_{x_1 \in S_1, x_2 \in S_2} |x_1 - x_2|. \tag{8}$$

Remark: The reason for introducing this "distance function" is that we think of H as defining the Hamiltonian for a one-dimensional, finite- range quantum system, with different "sites" of the system corresponding to different eigenvectors of B, and then the distance function is the distance between different sets of sites.

Further, we say that a vector w is "supported on set S for position operator B" if w is a linear combination of eigenvectors of B whose corresponding eigenvalues are in set S. Finally, for any set S we define the projector $P(S, B)$ to be the projector onto eigenvectors of B whose corresponding eigenvalues lie in set S. We now give the Lieb-Robinson bound:

Theorem 4. *Let H have the properties*

1*: $\|H\| \leq 1$.*

2*: For any two vectors v_1, v_2 which are eigenvectors of B with corresponding eigenvalues x_1, x_2, and with $|x_1 - x_2| \geq \Delta$, we have $(v_1, H v_2) = 0$.*

Define

$$v_{LR} = e^2 \Delta. \tag{9}$$

Then, for any vector v supported on a set S_1 for position operator B, and for any projector $P(S_2, B)$, we have

$$|P(S_2, B) \exp(-iHt) v| \leq e^{-\mathrm{dist}(S_1,S_2)/\Delta} |v| \tag{10}$$

for any

$$|t| \leq \mathrm{dist}(S_1, S_2)/v_{LR}. \tag{11}$$

Proof. Expand $\exp(-iHt)v$ in a power series as $v - iHtv - (H^2/2)t^2 v + \ldots$. Then, by assumption, $P(S_2, B)(-it)^n (H^n/n!)v$ vanishes for $n < \mathrm{dist}(S_1, S_2)/\Delta$. Let $m = \lceil \mathrm{dist}(S_1, S_2)/\Delta \rceil$. Then,

$$\begin{aligned} |\sum_{n \geq m} (-it)^n (H^n/n!) v| &\leq \sum_{n \geq m} (|t|^n/n!)|v| \\ &\leq \frac{1}{e} \sum_{n \geq m} (e|t|/n)^n |v| \\ &\leq \frac{1}{e} (e|t|/m)^m \frac{1}{1 - e|t|/m} |v|. \end{aligned} \tag{12}$$

For the given v_{LR} and t, the result follows. □

Remark: the proof of this Lieb-Robinson bound is significantly simpler than the proofs of the corresponding bounds for many-body systems considered elsewhere. The power series technique used here does not work for such systems.

B. Construction of X

In this subsection, we construct the matrix X from B. We define a function $Q(x)$ by

$$Q(x) = \Delta \lfloor x/\Delta + 1/2 \rfloor. \tag{13}$$

Then, we set

$$X = Q(B). \tag{14}$$

Note that $|Q(x) - x| \leq \Delta/2$ for all x, and $Q(x)/\Delta$ is always an integer. Then,

$$\|X - B\| \leq \epsilon_2, \tag{15}$$

with

$$\epsilon_2 = \Delta/2. \tag{16}$$

By (**2**) in lemma (1), the matrix H is a block tridiagonal matrix when written in a basis of eigenstates of X, with eigenvalues of X ordered in increasing order.

III. CONSTRUCTION OF NEW BASIS

In this section we construct the basis to make H close to a block diagonal matrix and X close to a block identity matrix. This completes step (**3**) of the construction of A' and B'. We refer to the basis constructed in this step as the "new basis" and we refer to the basis in which X is diagonal as the "old basis".

There will be a total of $n_{cut} + 1$ different blocks in the new basis, where n_{cut} is chosen later. Before constructing the new basis, we give some definitions. We define an interval I_i by $I_i = [-1 + 2(i-1)/n_{cut}, -1 + 2i/n_{cut})$ for $1 \leq i < n_{cut}$ and $I_i = [-1 + 2(i-1)/n_{cut}, -1 + 2i/n_{cut}]$ for $i = n_{cut}$. Let J_i be the matrix given by projecting H onto the subspace of eigenvalues of X lying in this interval I_i, and call this subspace $\mathcal{B}_i$. Then, in the old basis of eigenvalues of X, J_i is block tridiagonal with at least L different blocks, where $L = \lfloor (2/n_{cut}\Delta) - 1 \rfloor$ (some of these blocks might have dimension zero if B happens to have fewer than L distinct eigenvalues in that interval). We will choose n_{cut} later so that $L >> 1$ and so the new basis will have fewer blocks than the old basis. Before constructing the new basis we need the following lemma.

We claim that:

Lemma 2. *Let J be a Hermitian block tridiagonal matrix, with $\|J\| \leq 1$, acting on a space $\mathcal{B}$. Let there be L blocks, so that the space $\mathcal{B}$ has L orthogonal subspaces, which we write $\mathcal{V}_j$ for $j = 1, ..., L$, with $(v, Jw) = 0$ for $v \in \mathcal{V}_i$ and $w \in \mathcal{V}_j$ with $|i - j| > 1$. Then, there exists a space $\mathcal{W}$ which is a subspace of $\mathcal{B}$ with the following properties:*

- **1**: *The projection of any normalized vector $v \in \mathcal{V}_1$ onto the orthogonal complement of $\mathcal{W}$ has norm bounded by ϵ_3 where ϵ_3 is equal to $1/L^{1/3}$ times a function growing slower than any power of L.*
- **2**: *For any normalized vector $w \in \mathcal{W}$, the projection of Jw onto the orthogonal complement of $\mathcal{W}$ has norm bounded by ϵ_4, where ϵ_4 is equal to $1/L^{1/3}$ times a function growing slower than any power of L.*
- **3**: *The projection of any normalized vector $v \in \mathcal{V}_L$ onto $\mathcal{W}$ has norm bounded by ϵ_5, where ϵ_5 is equal to a function decaying faster than any power of L.*

Proof. This lemma is the key step in the proof of the main theorem, and the proof of this lemma is given in the next section. □

For each i, $1 \leq i \leq n_{cut}$, we apply lemma (2) to the matrix $J = J_i$ defined on the space $\mathcal{B} = \mathcal{B}_i$. For given i, we refer to the space $\mathcal{W}$ as constructed in lemma (2) as $\mathcal{W}_i$ and we refer to the subspaces $\mathcal{V}_j$ defined in lemma (2) as $\mathcal{V}_j(i)$. Let $\mathcal{B}_i$ have dimension $D_B(i)$ and let $\mathcal{W}_i$ have dimension $D_W(i)$. Let $\mathcal{W}_i^{\perp}$ denote the $D_B(i) - D_W(i)$-dimensional space which is the orthogonal complement of $\mathcal{W}_i$. Let $\mathcal{V}_j(i)$ have dimension $d_j(i)$. By properties **(1,2)** in lemma (2), $D_B(i) \geq d_1(i)$ and $D_B(i) \leq D_W(i) - d_L(i)$.

The new basis has $n_{cut} + 1$ blocks, which we label by $i = 0, 1, ..., n_{cut}$. For $1 \leq i < n_{cut}$, we define the i-th block of the new basis to be the space spanned by $\mathcal{W}_{i+1}$ and $\mathcal{W}_i^{\perp}$. For $i = 0$, the i-th block is the space $\mathcal{W}_{i+1} = \mathcal{W}_1$. For $i = n_{cut}$, the i-th block is the space $\mathcal{W}_i^{\perp} = \mathcal{W}_{n_{cut}}^{\perp}$.

Then, the matrix H is a block band matrix in this new basis. The matrix H will have terms on the main diagonal, on the diagonals directly above and below the main diagonal, and on the diagonals above and below those, so it only has terms within distance two of the main diagonal. The block-off-diagonal terms above and below the main diagonal arise from three sources. First, the matrix J_i contains non-vanishing matrix elements between the spaces $\mathcal{W}_i$ and $\mathcal{W}_i^{\perp}$, and those spaces are now in different blocks. However, by property (**2**) in lemma (2), these matrix elements are bounded by ϵ_4. Second, there are non-vanishing matrix elements between the subspace $\mathcal{W}_{i-1}^{\perp}$ and $\mathcal{V}_1^i$, and $\mathcal{V}_1^i$ may not be completely contained in subspace $\mathcal{W}_i$. However, by property **(1)** in lemma (2), these contribute only ϵ_3 to the norm of the block-off-diagonal terms of H in the new basis. Third, there are non-vanishing matrix elements between $\mathcal{W}_i$ and $\mathcal{V}_L^i$, and $\mathcal{V}_L^i$ may not be completely contained in subspace $\mathcal{W}_i^{\perp}$. However, by property **(1)** in lemma

(2), these contribute only ϵ_5 to the norm of the block-off-diagonal terms of H in the new basis. Therefore, these block-off-diagonal terms in H are bounded in operator norm by

$$2(\epsilon_3+\epsilon_4+\epsilon_5). \tag{17}$$

The terms which are on the diagonals two above or two below the main diagonal arise from the fact that there are non-vanishing matrix elements between $\mathcal{V}_L^i$ and $\mathcal{V}_1^{i+1}$, and $\mathcal{V}_L^i$ has some non-vanishing overlap with $\mathcal{W}_i$ and $\mathcal{V}_1^{i+1}$ has some non-vanishing overlap with $\mathcal{W}_{i+1}^{\perp}$. These terms are bounded by $\epsilon_3\epsilon_5$.

Define B' to be the block identity matrix (in the new basis) which is equal to $-1+2i/n_{cut}$ times the identity matrix in the i-th block. Since each block i in the new basis lies within the space spanned by $\mathcal{B}_i$ and $\mathcal{B}_{i+1}$ we have

$$\|B'-B\|\leq 2/n_{cut}. \tag{18}$$

Remark: Here is a sketch of the above procedure, in a case where H has 8 blocks and $n_{cut}=2$. The matrix originally looks like

$$\begin{pmatrix}
... & ... & & & & & & \\
... & ... & ... & & & & & \\
 & ... & ... & ... & & & & \\
 & & ... & ... & ... & & & \\
 & & & ... & ... & ... & & \\
 & & & & ... & ... & ... & \\
 & & & & & ... & ... & ... \\
 & & & & & & ... & ...
\end{pmatrix} \tag{19}$$

where the ... indicate non-vanishing entries. We combine the entries in the first 4 blocks into a matrix J_1 and the entries in the last 4 into a matrix J_2 so H looks like

$$\begin{pmatrix} J_1 & ... \\ ... & J_2, \end{pmatrix} \tag{20}$$

where the ... couples only the L-th block of space $\mathcal{B}_1$ to the 1-st block of space $\mathcal{B}_2$. Then, we apply lemma 2 to decompose $\mathcal{B}_1$ into spaces $\mathcal{W}_1$ and $\mathcal{W}_1^{\perp}$ so that J_1 looks like

$$\begin{pmatrix} ... & O(\epsilon_4) \\ O(\epsilon_4) & ... \end{pmatrix} \tag{21}$$

and similarly for J_2. Inserting Eq. (21) into Eq. (20), H looks like (in the new basis)

$$\begin{pmatrix}
... & O(\epsilon_4) & O(\epsilon_5) & O(\epsilon_3\epsilon_5) \\
O(\epsilon_4) & ... & ... & O(\epsilon_3) \\
O(\epsilon_5) & ... & ... & O(\epsilon_4) \\
O(\epsilon_3\epsilon_5) & O(\epsilon_3) & O(\epsilon_4) & ...
\end{pmatrix} \tag{22}$$

which is close to the block diagonal matrix,

$$\begin{pmatrix}
... & & & \\
 & ... & ... & \\
 & ... & ... & \\
 & & & ...
\end{pmatrix} \tag{23}$$

which has $3=n_{cut}+1$ blocks.

IV. PROOF OF LEMMA 2

Let the space $\mathcal{V}_1$ be d_1 dimensional, with orthonormal basis vectors $v_1,...,v_{d_1}$. Let S denote the D_B-by-d_1 matrix whose columns are these basis vectors, so that S is an isometry.

Define a function $\mathcal{F}(\omega_0,r,w,\omega)$ as follows. Let $\mathcal{F}(0,0,1,\omega)=1$ for $\omega=0$. Let $\mathcal{F}(0,0,1,\omega)=0$ for $|\omega|\geq 1$. Let $\mathcal{F}(0,0,1,\omega)=\mathcal{F}(0,0,1,-\omega)$. For $0\leq\omega\leq 1$, choose $\mathcal{F}(0,0,1,\omega)$ to be infinitely differentiable so that the

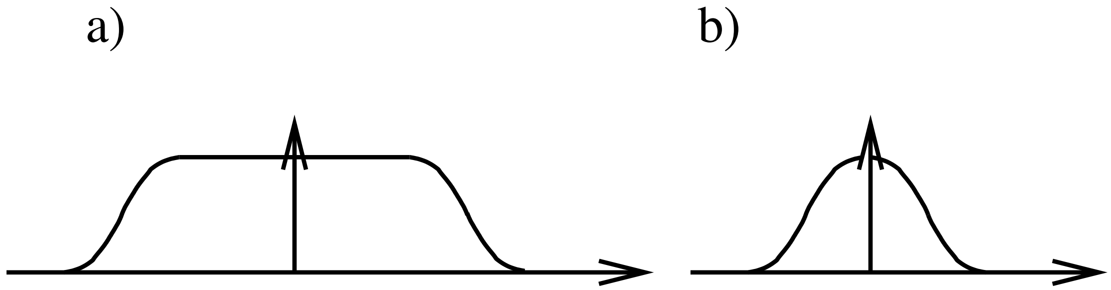

FIG. 1: Sketch of (a) $\mathcal{F}(0,1,1,\omega)=\mathcal{F}(-1,0,1,\omega)+\mathcal{F}(0,0,\omega)+\mathcal{F}(1,0,1,\omega)$ and (b) $\mathcal{F}(0,0,1,\omega)$.

Fourier transform of $\mathcal{F}(0,0,1,\omega)$, which we write $\tilde{\mathcal{F}}(0,0,1,t)$, is bounded by a function which decays faster than any polynomial. Finally, we impose $\mathcal{F}(0,0,1,\omega)+\mathcal{F}(0,0,1,1-\omega)=1$ for $0\leq\omega\leq 1$.

For general ω_0,r,w, define the function $\mathcal{F}(\omega_0,r,w,\omega)$ by $\mathcal{F}(\omega_0,r,w,\omega)=1$ for $|\omega-\omega_0|\leq r$, and $\mathcal{F}(\omega_0,r,w,\omega)=\mathcal{F}(0,0,1,(|\omega-\omega_0|-r)/w)$ for $|\omega-\omega_0|\geq r$. Then $\mathcal{F}(\omega_0,r,w,\omega)=0$ for $|\omega-\omega_0|\geq r+w$. For $r\geq 0$ and $w>0$, the function $\mathcal{F}(\omega_0,r,w,\omega)$ is infinitely differentiable with respect to ω everywhere. The functions $\mathcal{F}(0,1,1,\omega)$ and $\mathcal{F}(0,0,1,\omega)$ are sketched in Fig. 1a,b; the variable r denotes the width of the flat part at the center of the function, while w denotes the width of the changing part of the function. Since $\mathcal{F}(0,0,1,\omega)$ is infinitely differentiable, there is a function $T(x)$ which decays faster than any polynomial such that:

$$\int_{|t|\geq t_0}\mathrm{d}t|\tilde{\mathcal{F}}(\omega_0,w,w,t)|\leq T(wt_0),\tag{24}$$
$$\int_{|t|\geq t_0}\mathrm{d}t|\tilde{\mathcal{F}}(\omega_0,0,w,t)|\leq T(wt_0).$$

The operator norm of J is bounded by 1. The idea of the proof is to divide the interval of eigenvalues of J, which is $[-1,1]$, into various small overlapping windows. Then, for each interval centered on a frequency ω, we will construct vectors given by approximately projecting vectors in $\mathcal{V}_1$ onto the space spanned by eigenvectors of J with eigenvalues lying in that interval; we call the spaces of these vectors $\mathcal{X}_i$, where i labels the particular window. Then, each of these projected vectors x will have the property that Jx is close to ωx. This will be the key step in ensuring property **(2)** in the claims of the lemma. The idea of *approximate* projection is important here. In fact, we will use the smooth filter functions $\mathcal{F}(\omega_0,r,w,\omega)$ above. The smoothness will be essential to ensure that the vectors x have most of their amplitude in the first blocks rather than the last blocks. Since the vectors in the spaces $\mathcal{X}_i$ are approximate projections of vectors in $\mathcal{V}_1$ into different windows, we will be able to approximate any vector $v_1\in\mathcal{V}_1$ by a vector in the space spanned by the $\mathcal{X}_i$ simply by adding up the projections of v_1 in each different window. Because the windows overlap, the vectors may not be orthogonal to each other; the overlap between vectors is something we will need to bound (see Eq. (49) below). To control the overlap, we choose $\mathcal{W}$ to be a subpace of the space spanned by the $\mathcal{X}_i$ as explained below; this will then require us to be careful to ensure that we are still able to approximate vectors in $\mathcal{V}_1$ by vectors in $\mathcal{W}$.

Let n_{win} be some even integer chosen later. We will choose

$$n_{win}=L/F(L),\tag{25}$$

where the function $F(L)$ is a function that grows slower than any power of L and is defined further below. The choice of function $F(L)$ will depend only on the function $T(x)$ defined above.

For each $i=0,...,n_{win}-1$, define

$$\omega(i)=-1+2i/(n_{win}-1).\tag{26}$$

Define

$$\kappa=2/(n_{win}-1),\tag{27}$$

so $\omega(i)=-1+i\kappa$.

When $\omega(i)$ and κ are chosen as above, we have $\sum_{i=0}^{n_{win}-1}\mathcal{F}(\omega(i),0,\kappa,\omega)=1$ for $-1\leq\omega\leq 1$. See Fig. 2(a) to see a sketch of three functions $\mathcal{F}(\omega(i-1),0,\kappa,\omega)$, $\mathcal{F}(\omega(i),0,\kappa,\omega)$, and $\mathcal{F}(\omega(i+1),0,\kappa,\omega)$; as $\mathcal{F}(\omega(i),0,\kappa,\omega)$ decreases for $\omega(i)\leq\omega\leq\omega(i+1)$, the function $\mathcal{F}(\omega(i+1),0,\kappa,\omega)$ is increasing to keep the sum constant.

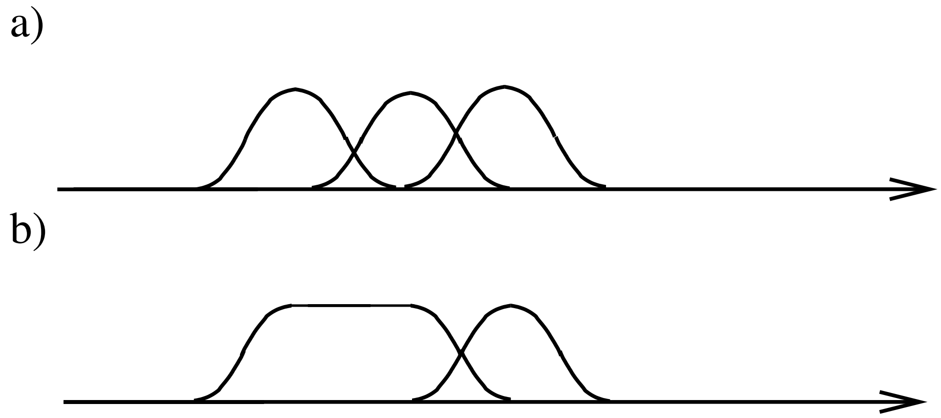

FIG. 2: a) Sketch of overlapping windows. b) Re-arrangement of windows as discussed in section on tridiagonal matrices.

A. Construction of Spaces $\mathcal{X}_i$

To construct $\mathcal{X}_i$, we define the matrix τ_i by

$$\tau_i \; = \; \mathcal{F}(\omega(i), 0, \kappa, J)S. \tag{28}$$

Define

$$\lambda_{min} = 1/(n_{win}L^2). \tag{29}$$

Compute the eigenvectors of the matrix $\tau_i^\dagger \tau_i$. For each eigenvector x_a with eigenvalue greater than or equal to λ_{min} compute $y_a = \tau_i x_a$. Let $\mathcal{X}_i$ be the space spanned by all such vectors y_a. Let Z_i project onto the eigenvectors x_a with eigenvalue less than λ_{min}; the projector Z_i will be used later in computing the error estimates.

Remark: To understand this construction, in Fig. 2a we sketch the functions $\mathcal{F}(\omega(i-1), 0, \kappa, \omega)$, $\mathcal{F}(\omega(i), 0, \kappa, \omega)$, and $\mathcal{F}(\omega(i+1), 0, \kappa, \omega)$, which form partially overlapping windows. Note that the vectors $\mathcal{F}(\omega(i), 0, \kappa, J)Sx_1$ and $\mathcal{F}(\omega(i \pm 1), 0, \kappa, J)Sx_2$, for arbitrary x_1, x_2, need not be orthogonal.

B. Properties of $\mathcal{X}_i$

This subsection establishes certain properties of the $\mathcal{X}_i$. It is primarily intended to motivate the construction thus far. We will show that the $\mathcal{X}_i$ have three properties which are closely related to the three properties we desire to show in lemma (2).

First, for any normalized vector $v \in \mathcal{V}_1$, the projection of v onto the orthogonal complement of the space spanned by the $\mathcal{X}_i$ is bounded by $\sqrt{2n_{win}\lambda_{min}} = \sqrt{2}/L$. To show this, for any $v \in \mathcal{V}_1$, with $|v| = 1$, we write $v = Sx$ with $|x| = 1$, and then

$$\begin{aligned} |v - \sum_{i=0}^{n_{win}-1} \tau_i(1-Z_i)x|^2 \; &= \; |\sum_{i=0}^{n_{win}-1} \tau_i Z_i x|^2 \\ &\le \; 2\sum_{i=0}^{n_{win}-1} |\tau_i Z_i x|^2 \\ &\le \; 2n_{win}\lambda_{min} \\ &\le \; 2/L^2. \end{aligned} \tag{30}$$

The factor of 2 in the first inequality follows because $(\tau_i Z_i x, \tau_j Z_j x) = 0$ for $|i-j| > 1$, but may be non-vanishing for $i = j \pm 1$. Similar factors of 2 occur in several other places.

Second, each space $\mathcal{X}_i$ is an approximate eigenspace of J. That is, for any $v_i \in \mathcal{X}_i$, we have

$$|(J - \omega(i))v_i| \le \kappa |v_i|. \tag{31}$$

Third, for any vector $y \in \mathcal{X}_i$, the norm of the projection of y onto $\mathcal{V}_L$ is bounded by $|y|$ times a function growing slower than any power of L. It is here that we will pick the function $F(L)$ and use the Lieb-Robinson bounds. Let $\tilde{\mathcal{F}}(\omega_0, r, w, t)$ denote the Fourier transform of $\mathcal{F}(\omega_0, r, w, \omega)$ with respect to the last variable ω. Then, for any x with $(1 - Z_i)x = x$, we find that $y = \tau_i x = \mathcal{F}(\omega(i), 0, \kappa, J)Sx$ is equal to

$$y = \int \mathrm{d}t\tilde{\mathcal{F}}(\omega(i), 0, \kappa, t)\exp(iJt)Sx. \tag{32}$$

We use the Lieb-Robinson bounds for matrix J, by defining a position matrix which is equal to i in the i-th block. Using the Lieb-Robinson bounds, for time $t \leq L/v_{LR}$, with $v_{LR} = e^2$, we find that the norm of the projection of $\exp(iJt)Sx$ onto the space $\mathcal{V}_L$ is bounded by $\exp(-L)$. At the same time, the integral $\int_{|t| \geq L/v_{LR}} \mathrm{d}t\tilde{\mathcal{F}}(\omega(i), 0, \kappa, t)$ is bounded by $T(2L/v_{LR}n_{win}) = T(2F(L)/e^2)$. Since $T(x)$ decays faster than any negative power of x, we can choose an $F(x)$ which grows slower than any power of x such that $T(2F(L)/e^2)$ still decays faster than any negative power of L. Thus, since $|y| \geq \lambda_{min}|x|$ by construction, for this choice of $F(x)$ the norm of the projection of any vector $y \in \mathcal{W}_i$ onto $\mathcal{V}_L$ is bounded by $|y|$ times a function decaying faster than any negative power of L.

The reason for picking $\lambda_{min} > 0$ is to help establish the third property above. Let us give an example of a situation where we would encounter problems if we have taken $\lambda_{min} = 0$. Consider a matrix of the form

$$\begin{pmatrix} 0 & 1/4 & & & & & \\ 1/4 & 0 & 1/4 & & & & \\ & 1/4 & 0 & 1/4 & & & \\ & & 1/4 & 0 & 1/4 & & \\ & & & ... & & & \\ & & & & 1/4 & 0 & 1/4 \\ & & & & & 1/4 & 1/2 \end{pmatrix} \tag{33}$$

Here, each block has size one. If it weren't for the "1/2" in the last line, this matrix would have operator norm slightly less than 1/2. However, because of the 1/2, this matrix has one eigenvalue greater than 1/2. For this particular choice of matrix, this eigenvalue is close to 5/8. The corresponding eigenvector is localized near the last block, and is exponentially small in the first block. If we project a vector in $\mathcal{V}_1$ into a narrow window centered on $\omega(i) = 5/8$, the result will project onto this eigenvector, and thus the resulting state will have large amplitude on $\mathcal{V}_L$. However, for such a window, we would find that τ_i would be exponentially small, and so we would not include this vector in $\mathcal{X}_i$.

The properties we have established for spaces $\mathcal{X}_i$ are closely related to the properties in lemma (2) that we are trying to establish. Unfortunately, the spaces $\mathcal{X}_i$ need not be orthogonal, and in fact may be very far from orthogonal. This can lead to problems like the following: suppose we have two vectors, $v_1 \in \mathcal{X}_1$ and $v_2 \in \mathcal{X}_2$. We know that the projection of v_1 onto $\mathcal{V}_L$ is small compared to $|v_1|$, and we know the same thing for v_2; however, we don't know that the projection of $v_1 + v_2$ onto $\mathcal{V}_L$ is small compared to $|v_1 + v_2|$ because we don't know how $|v_1 + v_2|$ compares to $|v_1|$ and $|v_2|$. We have two different ways of dealing with this: in the next subsection, we present a construction for block tridiagonal matrices that involves choosing a subspace of the space spanned by the $\mathcal{X}_i$. In a later section on tridiagonal matrices, we present a much simpler construction that involves combining several windows into one; the reader may prefer to read that section first.

C. Construction of $\mathcal{W}$

We now construct the space $\mathcal{W}$. Let each space $\mathcal{X}_i$ have dimension D_i. In each space $\mathcal{X}_i$ we can find an orthonormal basis of vectors, $v_{i,b}$, for $b = 1, ..., D_i$. We define a block tridiagonal matrix ρ of inner products of vectors $v_{i,b}$ as follows: the i-th block (for $0 \leq i < n_{win}$) has dimension D_i, and on the diagonal the matrix is equal to the identity matrix. Above the diagonal, the block in the i-th row and $i+1$-st column is equal to the matrix of inner products $(v_{i,b}, v_{i+1,c})$ for $b = 1, ..., D_i$ and $c = 1, ..., D_{i+1}$. Note that for $|i - j| > 1$, the spaces $\mathcal{X}_i$ and $\mathcal{X}_j$ are orthogonal, so that the matrix ρ is block tridiagonal. We define a new vector space $\mathcal{R}$ to be a space of dimension $\sum_{i=0}^{n_{win}-1} D_i$. The matrix ρ is Hermitian and positive semidefinite. It is equal to $\rho = A^\dagger A$, for some matrix A which is also block tridiagonal. The matrix A is a linear operator from $\mathcal{R}$ to $\mathcal{B}$; it is simply a matrix whose columns, in a given block, are different basis vectors for the space $\mathcal{X}_i$ corresponding to that block.

Remark: The matrix ρ is block tridiagonal. To motivate what follows, consider the following circular reasoning: given that ρ is block tridiagonal, if we knew that theorem (2) were true, we could find a basis in which ρ was approximately diagonal and in which a position operator, a block diagonal matrix equal to i/n_{win} in the i-th block, was also approximately diagonal. Then we choose $\mathcal{W}$ to be the space spanned by vectors of the form Aw_i, where w_i are basis vectors in this basis for which the diagonal entry of ρ are not too close to zero (how close is something we

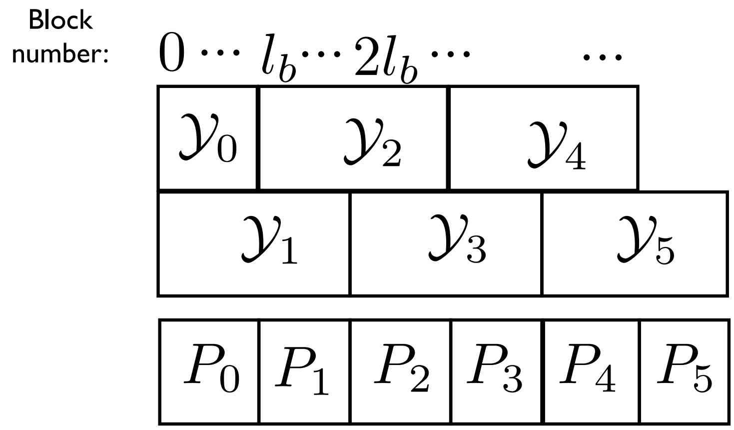

FIG. 3: Sketch of which blocks are in which subspaces $\mathcal{Y}_i$ for $n_{sb} = 6$, as well as which blocks are in the range of the P_i.

would pick later). Then, we would know that the vectors Aw_i and Aw_j are not degenerate for $i \neq j$, and the operator A would be an approximate isometry from the space spanned by the w_i to $\mathcal{W}$. Also, we would find that Aw_i was an approximate eigenvector of J. We would know that any vector $v \in \mathcal{V}_1$ had small projection orthogonal to $\mathcal{W}$, since v could be written as Sx and, while x may have some projection onto vectors w_i for which the diagonal entry of ρ is very close to zero, the error in v we make by dropping those vectors from x is small. This would give the space $\mathcal{W}$ the properties we are trying to construct. Unfortunately, of course, we are trying to prove theorem (2), so this line of reasoning does not help. However, we do not need such a strong result in the present construction as will be seen below. Importantly, if there is a vector w such that $(w, \rho w)$ is small, it leads to only a small error in our ability to approximate vector in $v \in \mathcal{V}_1$ if the vector w is orthogonal to the space $\mathcal{W}$. We will also make use of a related fact: if there is a vector $w = w_1 + w_2$ such that $(w_1 + w_2, \rho(w_1 + w_2))$ is small, then this means that Aw_1 is close to $-Aw_2$. Suppose $Aw_1 \in \mathcal{X}_1$ and $Aw_2 \in \mathcal{X}_2$. Then, we can take $\mathcal{W}$ to be the space spanned by $\mathcal{X}_2, \mathcal{X}_3, ...$ and spanned by the subspace of $\mathcal{X}_1$ orthogonal to Aw_1, and this leads to only a small error in our ability to approximate vectors $v \in \mathcal{V}_1$ by vectors in $\mathcal{W}$. This is the basic idea behind the construction that follows.

We define spaces $\mathcal{Y}_i$, for $i = 0, ..., n_{sb} - 1$, as follows, where n_{sb} is the smallest even integer greater than or equal to n_{win}/l_b with the "block length" l_b being an integer equal to

$$l_b = \lfloor n_{win}^{1/3} \rfloor. \tag{34}$$

Here, "sb" stands for "super-block" as we combine several blocks into one superblock. We pick $\mathcal{Y}_i$ to be the subspace of $\mathcal{R}$ spanned by the vectors in blocks from the $(i-1)l_b$-th block to the $(i+1)l_b - 1$-th block. That is, it is the subspace spanned by vectors in blocks $(i-1)l_b, (i-1)l_b + 1, (i-1)l_b + 2, ..., (i+1)l_b - 1$. Therefore, $\mathcal{Y}_i$ is orthogonal to $\mathcal{Y}_j$ for $|i - j| > 1$. The space spanned by the $\mathcal{X}_i$ for $i = 0, ..., n_{win} - 1$ is the same as the space spanned by the $A\mathcal{Y}_i$ for $i = 0, ..., n_{sb} - 1$; we will choose the space $\mathcal{W}$ to be a subspace of this space. Let P_i project onto the subspace of $\mathcal{R}$ spanned by the blocks from the il_b-th block to the $(i+1)l_b - 1$-th block. For notational convenience later (and to avoid various off-by-one errors), we define $P_{-1} = 0$, and we define $\mathcal{X}_i$ for $i < 0$ to be the empty set.

In Fig. 3 we sketch the blocks used to define the spaces $\mathcal{Y}_i$ for the case $n_{sb} = 6$. The horizontal position in the figure indicates increasing block number, as marked in the top row. Space $\mathcal{Y}_i$ overlaps with space $\mathcal{Y}_{i\pm1}$, as seen. We have also sketched the range of the operators P_i.

We claim that

Lemma 3. *There exist spaces $\mathcal{N}_i$, for $i = 0, ..., n_{sb} - 1$ with the properties that:*

1: *$\mathcal{N}_i$ is a subspace of $\mathcal{Y}_i$.*

2: *For any vector $v \in \mathcal{N}_i$, the quantity $(v, \rho v)$ is bounded by $|v|^2/l_b^2$ times a function $F_0(l_b)$, which is growing slower than any power of l_b.*

3: *Let N_i project onto $\mathcal{N}_i$. For any vector v which is in the space spanned by eigenvectors of ρ with eigenvalue less than $1/l_b^2$, the sum $\sum_i |N_i v|^2$ is greater than or equal to $(1 - F_1(l_b))|v|^2$, where $F_1(l_b)$ is a function decaying faster than any negative power of l_b.*

4: *For any vector v which is in the space spanned by eigenvectors of ρ with eigenvalue less than $1/l_b^2$, we can find vectors $n_i \in \mathcal{N}_i$ such that*

$$v = \sum_i n_i + w^{\perp}, \tag{35}$$

with $w^{\perp}$ bounded by a quantity going to zero faster than any power of l_b, and $\mathrm{Re}((n_i, n_{i+1})) \geq -F_2(l_b)|v|^2$, where $F_2(l_b)$ goes to zero faster than any power of l_b and where $\mathrm{Re}(...)$ denotes the real part of a quantity.

Proof. Define a matrix M by

$$M = \begin{pmatrix} 0 & A \\ A^{\dagger} & 0 \end{pmatrix}. \tag{36}$$

Define a matrix O by

$$O = \int \mathrm{d}t \, \exp(iMt)\tilde{\mathcal{F}}(0, G(l_b)/l_b, G(l_b)/l_b, t), \tag{37}$$

where $G(l_b)$ is a function growing slower than any power of l_b to be chosen later. Note that since $\mathcal{F}$ is even in t, O is non-vanishing only in the upper left and lower right corners. We define O_0 to be the top left corner of O. For use later, define the matrix Q by

$$Q = \begin{pmatrix} 1 & 0 \end{pmatrix}. \tag{38}$$

For each i, define A_i to project onto the blocks with $(i-1/2)l_b \leq j < (i+1/2)l_b$. Compute the eigenvalues of $A_i O_0^2 A_i$. For each eigenvalue λ_a greater than or equal to λ_0, for some λ_0 chosen later, compute the corresponding eigenvector x. The quantity λ_0 will be chosen later to go to zero faster than any power of l_b. Let $\mathcal{N}_i$ be the space spanned by $(P_{i-1}+P_i)O_0 x$, for all such x. $\mathcal{N}_i$ is a subspace of $\mathcal{Y}_i$, as claimed.

Let v be any vector in $\mathcal{N}_i$. By definition, $v = (P_{i-1}+P_i)O_0 x$, for some x. By construction, $(O_0 x, \rho O_0 x)$ is bounded by $(2G(l_b)/l_b)^2$. We now use the Lieb-Robinson bounds for matrix M, by defining a position matrix which is equal to i in the i-th block. Using the Lieb-Robinson bounds, for time $t \leq l_b/2v_{LR}$, with $v_{LR} = e^2$, we find that the norm of $(1-P_{i-1}-P_i)Q\exp(iMt)Q^T x$, for x in the range of A_i, is bounded by $\exp(-l_b/2)$. At the same time, the integral $\int_{|t|\geq l_b/2v_{LR}} \mathrm{d}t\tilde{\mathcal{F}}(0, G(l_b)/l_b, G(l_b)/l_b, t)$ is bounded by $T(G(l_b)/2v_{LR})$. Since $T(x)$ decays faster than any negative power of x, we can choose a $G(x)$ which grows slower than any power of x such that $T(G(l_b)/2v_{LR})$ still decays faster than any negative power of l_b. Thus, since $|y| \geq \lambda_0 |x|$ by construction, we can find a λ_0 going to zero faster than any power of l_b such that $|(P_{i-1}+P_i)O_0 x - O_0 x|$ goes to zero faster than any power of l_b. Thus,

$$|(v,\rho v)| \;\leq\; |(O_0 x, \rho O_0 x)| + 2|(O_0 x, \rho(1-P_{i-1}-P_i)O_0 x)| + ((1-P_{i-1}-P_i)O_0 x, \rho(1-P_{i-1}-P_i)O_0 x). \tag{39}$$

Since ρ is bounded in operator norm, we have bounded all terms in the above equation. This verifies **2**.

Also, for any vector v which is a normalized eigenvector of ρ with eigenvalue less than $1/l_b^2$, we can write $v = \sum_i a_i v_i$, with v_i a normalized vector in the space projected onto by A_i and a_i a set of complex amplitudes such that $\sum_i |a_i|^2 = 1$. Then, since $O_0 v = v$, we have $\sum_i |a_i|^2 |(O_0 v_i, v)|^2 = |v|^2 = 1$. Decompose each v_i as $y_i + z_i$ where y_i is the projection of v_i onto eigenvectors of $A_i O_0^2 A_i$ with eigenvalue greater than or equal to λ_0 and z_i is the projection onto eigenvectors with eigenvalue less than λ_0. Note that $|O_0 z_i|$ is bounded by $\sqrt{\lambda_0}$. Thus, $|O_0 v_i - O_0 y_i|$ is bounded by $\sqrt{\lambda_0}$. Thus, since $\sum_i |a_i|^2 |(O_0 v_i, v)|^2 = 1$, we have $\sum_i |(O_0 a_i y_i, v)|^2 \geq 1 - 2\sqrt{\lambda_0}$. Further, using the Lieb-Robinson bounds, $|(P_{i-1}+P_i)O_0 y_i - O_0 y_i|$ is bounded by a quantity going to zero faster than any power of l_b. Thus, given the lower bound on $O_0 y_i$ that $|O_0 y_i| \geq \sqrt{\lambda_0}$, we can pick a λ_0 going to zero faster than any power of l_b such that $|(P_{i-1}+P_i)O_0 y_i - O_0 y_i|/|O_0 y_i|$ is bounded by a quantity going to zero faster than any power of l_b. Thus, the angle between $O_0 y_i$ and $|(P_{i-1}+P_i)O_0 y_i|$ goes to zero faster than any power of l_b, and so the projection of $O_0 y_i$ into N_i is equal to $O_0 y_i$ plus a vector going to zero faster than any power of l_b. Thus, $|N_i v|^2$ is greater than or equal to $|a_i|^2 |(O_0 y_i, v)|^2$ minus a quantity going to zero faster than any power of l_b. This verifies **3**.

The proof of **4** is similar to **3**. Write $v = \sum_i a_i v_i$ as in the above paragraph. As before, decompose $v_i = y_i + z_i$. Then, $|v - \sum_i (P_{i-1}+P_i)O_0 y_i|$ is bounded by a quantity going to zero faster than any power of l_b. Let $n_i = \sum_i (P_{i-1}+P_i)O_0 y_i$. Then, $n_i \in \mathcal{N}_i$ as claimed and we have upper bounded $|w^{\perp}| = |v - \sum_i n_i|$ as claimed. Now consider (n_i, n_{i+1}). This is equal to $(\sum_{j\leq i} n_j, \sum_{j>i} n_j)$. Note that $|\sum_{j\leq i} n_j| = |\sum_{j\leq i}(P_{j-1}+P_j)O_0 y_i|$. However, $|\Big(\sum_{j\leq i}(P_{j-1}+P_j)O_0 y_i\Big) - O_0 \sum_{j\leq i} v_i|$ is upper bounded by $|v|$ times a quantity going to zero faster than any power of l_b. Thus, since $\|O_0\| \leq 1$, we have upper bounded $|\sum_{j\leq i} n_j|$ by $|\sum_{j\leq i} v_j|$ plus $|v|$ times a quantity going to zero

faster than any power of l_b. Similarly, we have upper bounded $|\sum_{j>i} n_j|$ by $|\sum_{j>i} v_j|$ plus $|v|$ times a quantity going to zero faster than any power of l_b. Now consider $|\sum_j n_j|^2$. This is equal to

$$|\sum_j n_j|^2 = |\sum_{j\leq i} n_j|^2 + |\sum_{j>i} n_j|^2 + 2\mathrm{Re}((n_i, n_{i+1})). \tag{40}$$

We have upper bounded the first two terms on the right-hand side of the above equation by $|\sum_{j\leq i} v_j|^2 + |\sum_{j>i} v_j|^2$, plus $|v|^2$ times a quantity going to zero faster than any power of l_b. Since $|\sum_{j\leq i} v_j|^2 + |\sum_{j>i} v_j|^2 = |v|^2$, and since $|\sum_j n_j - v|$ is bounded by a quantity going to zero faster than any power of l_b so that $|\sum_j n_j|^2$ is superpolynomially close to $|v|^2$, we have lower bounded $\mathrm{Re}((n_i, n_{i+1}))$, as desired. □

Remark: The definition of M and O as block matrices in the above lemma is simply a trick to make the claims **2**, **3** in the lemma depend on l_b^{-2} rather than l_b^{-1} as we would have found without this trick of introducing block matrices. In physics jargon, near the edge of the band (eigenvalues close to zero for ρ which is a positive semi-definite matrix), we have dynamic critical exponent 2 rather than 1.

We now describe the construction. The numeric constant η below is some sufficiently small, positive, real number; the choice of this number will be discussed in lemma (4).

We iteratively construct a sequence of spaces $\mathcal{N}'_i$ for odd i which are subspaces of $\mathcal{N}_i$ as follows. For each $i = 1, 3, ...$, consider the space $\mathcal{N}_i$. Let Q_i be the projector onto the span of $\mathcal{N}_{i+1}, \mathcal{N}_{i-1}$ and $\mathcal{N}'_{i-2}$ (if $i = 1$, then $\mathcal{N}'-1$ is considered to be the empty space). Apply Jordan's lemma[28] to N_i and to Q_i to construct a complete orthonormal basis $n_{i,b}$ for $\mathcal{N}_i$ such that $(n_{i,b}, Q_i n_{i,c}) = 0$ for $b \neq c$. Let $\mathcal{N}'_i$ be the space spanned by vectors $n_{i,b}$ such that $|Q_i n_{i,b}|^2 \leq 1/2 + \eta$. Define $\mathcal{U}^\perp$ to be the space spanned by the $\mathcal{N}_i$ for even i and by the $\mathcal{N}'_i$ for odd i. Define $\mathcal{U}$ to be the subspace of $\mathcal{R}$ orthogonal to $\mathcal{U}^\perp$. We now define $\mathcal{W} = A\mathcal{U}$. Let P be the projector onto $\mathcal{W}$. We define U to be the projector onto $\mathcal{U}$. Thus, for any vector w with $Pw = w$, we have $w = Av$, for some v with $Uv = v$.

We claim one important property of this space $\mathcal{U}$:

Lemma 4. *Let η be a sufficiently small positive number. Let x_i be any vector in $\mathcal{Y}_i$. Consider the vector Ux_i. Project this vector into space $\mathcal{Y}_j$. The norm of the resulting vector is bounded by*

$$\text{const.} \times \exp(-|i-j|/\text{const.})|x_i|, \tag{41}$$

for some positive numeric constants.

Proof. We consider instead the vector $(1-U)x_i$ and bound the norm of the projection of that vector. Since the space $\mathcal{U}^\perp$ is the span of spaces $\mathcal{N}_j$ for j even and $\mathcal{N}'_j$ for j odd, the projection $(1-U)x_i$ can be computed by minimizing the quantity

$$|\sum_j a_j n_j - x_i|^2 \tag{42}$$

over all $|n_j|$ in $\mathcal{N}_j$ for j even and n_j in $\mathcal{N}'_j$ for j odd, with $|n_j| = 1$, and over all complex numbers a_j. Then, $(1-U)x_i = \sum_j a_j n_j$. For given x_i, let the minimum be obtained for some definite choice of vectors n_j. Then, we consider the minimum over a_j of (42). We can write Eq. (42) in a matrix form, by introducing a tridiagonal matrix M, with diagonal entries equal to unity and entries $M_{i,i+1} = (n_i, n_{i+1})$. Then, define a vector $\vec{x}$ which has its j-th entry equal to (n_j, x_i). Define a vector $\vec{a}$, with j-th entry equal to a_j. Then, the minimum over a_j is given by

$$\vec{a} = M^{-1}\vec{x}. \tag{43}$$

We will prove an exponential decay on entries of M^{-1}. That is, we will define $G = M^{-1}$ and prove that the matrix element G_{ij} decays exponentially in $|i-j|$. Then, since the only non-zero entries of $\vec{x}$ are the $i-1$,i, and $i+1$ entries, this will prove the desired result (41).

To prove this decay, it suffices to prove a lower bound on the smallest eigenvalue of M. By construction, we have the property that for odd i,

$$|M_{i,i+1}|^2 + |M_{i,i-1}|^2/(1-|M_{i-1,i-2}|^2) \leq (1/2) + \eta, \tag{44}$$

since $|M_{i,i-1}|^2/(1-|M_{i-1,i-2}|^2)$ is the projection of n_i onto the span of n_{i-1} and n_{i-2}. It suffices to prove the lower bound on the smallest eigenvalue for matrices M that saturate Eq. (44), that is, those for which the inequality becomes an equality. Further, it suffices to prove the lower bound for matrices the case $\eta = 0$, since any matrix that obeys Eq. (44) with non-zero η is close in operator norm (the distance depends on η) to a matrix that obeys Eq. (44)

with $\eta = 0$, and hence if there is a lower bound on the smallest eigenvalue for $\eta = 0$, then we have a lower bound on the smallest eigenvalue for sufficiently small η.

We prove the lower bound in the case $\eta = 0$ as follows. We choose λ to be some small number. We consider the matrix $M - \lambda I$, where I is the identity matrix, and let let $M_{(n)}$ denote a sub-matrix of $M - \lambda I$, containing the first n rows and columns, and define $G^{(n)}$ to be $M_{(n)}^{-1}$. Let G_n be the matrix element $G^{(n)}_{n,n}$. We will prove an upper bound on G_n for all n, for some non-zero value of λ, in the case $\eta = 0$. This will prove the lower bound.

To compute G_n, we have a recursion relation:

$$G_n = \frac{1}{1-\lambda-y^2} + \frac{x^2 G_{n-2} y^2/(1-\lambda-y^2)^2}{1 - x^2 G_{n-2}/(1-\lambda-y^2)}, \tag{45}$$

where $x = |M_{n-2,n-1}|$ and $y = |M_{n-1,n}|$.

We can apply this recursion relation twice to compute G_n from G_{n-4}. Let us call $a = |M_{n-4,n-3}|$, $b = |M_{n-3,n-2}|$, $x = |M_{n-2,n-1}|$, and $y = |M_{n-1,n}|$. Then, Eq. (44) implies that (taking $\eta = 0$)

$$\begin{aligned} a^2 + b^2 &\leq 1/2, \\ y^2 + x^2/(1-b^2) &\leq 1/2. \end{aligned} \tag{46}$$

For what follows, it suffices to consider the case in which the inequalities in the above equation become equalities:

$$\begin{aligned} a^2 + b^2 &= 1/2, \\ y^2 + x^2/(1-b^2) &= 1/2. \end{aligned} \tag{47}$$

We do an inductive proof of the bound on G_n. For use in the induction, we say that "property $*$ holds for G_n" if two conditions hold. First, $G_n \leq 2.1$ and second, if $G_n > 1.9$ then $|M_{n,n-1}| > 0.54$. The following claims were checked by computer assistance: we choose a fine grid on possible values of b, x (in this case, every possible value between 0 and 1 with a spacing of 10^{-4}). We then computed the values of a, y from Eq. (47), and verified that certain claims held; we then verified that the chosen spacing was sufficiently fine that other claims held for *all* possible choices of b, x.

We have chosen $\lambda = 0.02$, and have verified that, for *all* b, x, if $G_{n-4} \leq 1.9$, then $G_n \leq 1.9$ unless $y > 0.54$ (this was done by verifying a stronger claim for all b, x on the grid, namely that $G_n \leq 1.89$, and then using smoothness properties of the functions). We further verified that assuming $G_{n-4} \leq 1.9$, then the maximum possible value of G_n was at $x = 0$ and gave $2.08333... < 2.1$. Thus, we verified that if $G_{n-4} \leq 1.9$, then property $*$ holds for G_n. For use in the inductive result below, we call this result the "first simulation".

Next, we assumed that $b > 0.54$, and that $G_{n-4} \leq 2.1$, and again verified that $G_n \leq 1.9$ unless $y > 0.54$, and that the maximum possible value of G_n was at $x = 0$ and gave $2.08333... < 2.1$. That is, we verified that if $G_{n-4} \leq 2.1$ and $|M_{n-2,n-3}| > 0.54$ then property $*$ holds for G_n. Again, this was done by verifying stronger statements for all b, x on the grid and then using smoothness. We call the results in this paragraph the "second simulation".

Finally, we verified that if $G_n < 1.9$, then for any choice of $M_{n,n+1}$ and $M_{n+1,n+2}$ with $|M_{n,n+1}|^2 + |M_{n+1,n+2}|^2 \leq 1/2$, we have $G_{n+2} < 2.1$. This is the "third simulation".

Now we prove that G_m is bounded by 2.1 all even m. Start with G_0. Since $G_0 = 1$, we have that property $*$ holds for G_0. Now, consider any even m and assume that property $*$ holds for G_m. Also assume that $G_l < 2.1$ for all even $l < m$. Then, if $|M_{m,m-1}| > 0.54$, we use the inductive assumption that $G_{m-2} < 2.1$ and the second simulation above (taking $n = m + 2$ so we compute G_{m+2} from G_{m-2}, in which case the assumption $|M_{m,m-1}| > 0.54$ means that $b > 0.54$) to show that either $G_{m+2} < 1.9$ or $G_{m+2} < 2.1$ and $|M_{m+1,m+2}| > 0.54$. That is, if $|M_{m,m-1}| > 0.54$, we show that property $*$ holds for G_{m+2}. On the other hand, if $|M_{m,m-1}| \leq 0.54$, then $G_m < 1.9$. In this case, by the third simulation, $G_{m+2} < 2.1$, and by the first simulation (now taking $m = n$), either $G_{m+4} < 1.9$ or $G_{m+4} < 2.1$ and $|M_{m+3,m+4}| > 0.54$, so that property $*$ holds for G_{m+4}. Note that this induction is non-standard: we assume that property $*$ holds for G_m, and we prove that the same hypothesis holds for either $m + 2$ or for $m + 4$. However, we prove the bound that $G_m < 2.1$ for all even m. This bounds G_m for $\lambda = 0.02$; using monotonicity properties of the recurrence of G, we then bound it for all $\lambda \leq 0.02$. □

Remark: The existence of a lower bound on the smallest eigenvalue is not surprising given the following argument. The argument in this paragraph is intended as heuristic, but could perhaps be turned into an alternate proof of the lemma above. Consider weakening the conditions on the off-diagonal matrix M to the condition that $|M_{n-2,n-1}|^2 + |M_{n-1,n}|^2 \leq 1/2$ for even n. One may show in this case that G_n is bounded by 2 for all even n. To see this, assume that G_{n-2} is bounded by 2. Maximizing G_n over $M_{n-2,n-1}$ and $M_{n-1,n}$ subject to the constraint gives that G_n is bounded by 2. Thus, using induction G_n is bounded for even n. Now, G_n may become infinite for odd n in this case (suppose that $M_{0,1} = 0, M_{1,2} = 1/\sqrt{2}, M_{2,3} = 1/\sqrt{2}, M_{3,4} = 0$, so that G_3 is infinite). One may similarly show in

this fashion that M is positive semi-definite (simply repeat this inductive argument for the matrix $M + zI$ for any real $z > 0$ to show that G_n is bounded for all real $z > 0$). Now, in the lemma above we assumed stronger constraints on M; it is perhaps not surprising then that adding the stronger constraints means that rather than being merely positive semi-definite, M has a lower bound on its smallest eigenvalue.

D. Properties of $\mathcal{W}$

In this section we establish certain properties for the space $\mathcal{W}$. The main results are Eq. (49), controlling the overlap between vectors in this space, and Eq. (50), showing that for any vector v in the space spanned by $\mathcal{X}_i$ with $v = Ax$, the vector $Pv \in \mathcal{W}$ is close to v, where the maximum distance $|Pv - v|$ between the vectors depends on $|x|$.

First Property— By construction, for any vector $r \in \mathcal{U}$, with $|r| = 1$, we have

$$(r, \rho r) \geq \text{const.} \times (1/l_b^2), \tag{48}$$

for sufficiently large l_b.

To show Eq. (48), we bound the inner product between r and w for w in the span of eigenvectors of ρ with eigenvalue less than or equal to $1/l_b^2$. Any such w can be written as a linear combination of vectors w^{even} in the span of $\mathcal{N}_i$ for even i and w^{odd} in the span of $\mathcal{N}_i$ for odd i. Note that w^{even} is in $Q^\perp$. By **4** in lemma (3), the inner product (w^{even}, w^{odd}) is greater than or equal to minus a quantity going to zero faster than any power of l_b. Thus, $|(1-U)w|^2$ is greater than or equal to $|w^{even}|^2$ minus a quantity going to zero faster than any power of l_b.

Further, we write $w^{odd} = w_1 + w_3$ where $w_1 = \sum_{i=1,5,9,...} n_i$ and $w_3 = \sum_{3,7,11,...} n_i$, with $n_i \in \mathcal{N}_i$. Note that by construction, each n_i has projection at least $(1/2+\eta)|n_i|^2$ onto the span of spaces $\mathcal{N}'_{i-2}, \mathcal{N}_{i-1}, \mathcal{N}_{i+1}$. Further, this span of spaces $\mathcal{N}'_{i-2}, \mathcal{N}_{i-1}, \mathcal{N}_{i+1}$ is orthogonal to the span of $\mathcal{N}'_{j-2}, \mathcal{N}_{j-1}, \mathcal{N}_{j+1}$ if $|i-j| \geq 4$. Thus, the projection of w_1 onto the span of $\mathcal{N}_i$ over odd i and $\mathcal{N}'_i$ over even i is at least $(1/2+\eta)|w_1|^2$, and so $|(1-U)w_1|^2 \geq (1/2+\eta)|w_1|^2$ and similarly $|(1-U)w_3|^2 \geq (1/2+\eta)|w_3|^2$. Since $(w_1, w_3) = 0$, we have $|(1-U)(w_1+w_3)|^2 \geq |(1-U)w_1|^2 + |(1-U)w_3|^2 - 2|(w_1, Uw_3)| \geq |(1-U)w_1|^2 - |(1-U)w_3|^2 - 2|Uw_1||Uw_3|$. Since $2|Uw_1||Uw_3| \leq |Uw_1|^2 + |Uw_3|^2$, $|(1-U)(w_1+w_3)|^2 \geq |(1-U)w_1|^2 + |(1-U)w_3|^2 - |Uw_1|^2 - |Uw_3|^2 \geq 2\eta(|w_1|^2 + |w_3|^2)$ and so $|(1-U)w|^2 \geq 2\eta|w|^2$ minus a quantity going to zero superpolynomially in l_b. Therefore, having upper bounded $|Uw|$, we have upper bounded (r, w), so Eq. (48) follows.

For any $v \in \mathcal{W}$, we can find $x_i \in \mathcal{X}_i$ for $i = 0, ..., n_{win} - 1$, such that $v = \sum_i x_i$. Therefore, from Eq. (48), for any $v \in \mathcal{W}$, we can find x_i, $i = 0, ..., n_{win} - 1$ with $x_i \in \mathcal{X}_i$ and $v = \sum_i x_i$ with

$$|v|^2 \geq \text{const.} \times (1/l_b)^2 \sum_{i=0}^{n_{win}-1} |x_i|^2. \tag{49}$$

Second Property— We also claim that for any vector v in the space spanned by the $\mathcal{X}_i$, such that $v = Ax$ that

$$|Pv - v| \leq \text{const.} \times (\sqrt{F_0(l_b)}/l_b)|x|. \tag{50}$$

To show Eq. (50), any vector x can be written as a linear combination of a vector in $\mathcal{Q}$ and a vector in $x^\perp \in \mathcal{Q}^\perp$. Let $x = \sum_i x_i$, with $x_i \in \mathcal{Y}_i$ and $\sum_i |x_i|^2 = |x|^2$. The vector $x^\perp = (1-U)x$. Let $x^\perp = \sum_j a_j n_j$, with n_j in $\mathcal{N}_j$ for j even and n_j in $\mathcal{N}'_j$ for j odd as in lemma (4). We bound $|Ax^\perp|^2$ by const. $\times (F_0(l_b)/l_b^2) \sum_{|j-k|\leq 1} |n_j||n_k| \leq$ const. $\times (F_0(l_b)/l_b^2) \sum_j |n_j|^2 \leq$ const. $\times (F_0(l_b)/l_b^2) \sum_j |x_j|^2$, where the last inequality uses the exponential decay on matrix element of G from lemma (4).

Any vector $v = \sum_{i=0}^{n_{win}-1} x_i$ with $x_i \in \mathcal{X}_i$ can be written as $v = Ax$ with $|x|^2 = \sum_{i=0}^{n_{win}-1} |x_i|^2$. Therefore, Eq. (50) implies that for any $v = \sum_{i=0}^{n_{win}-1} x_i$, with $x_i \in \mathcal{X}_i$, we have

$$|Pv - v| \leq \text{const.} \times (\sqrt{F_0(l_b)}/l_b)\sqrt{\sum_{i=0}^{n_{win}-1} |x_i|^2}. \tag{51}$$

E. Verification of Claims

We now verify the claims regarding the subspace $\mathcal{W}$.

Proof of First Claim— To prove (**1**), note that for any vector $v \in \mathcal{B}$ we have

$$v \;=\; \sum_{i=0}^{n_{win}-1} \mathcal{F}(\omega(i), 0, 2n_{win}, J)v. \tag{52}$$

For any $v \in \mathcal{V}_1$, with $|v| = 1$, we can write $v = Sx$ with $|x| = 1$, and then, from Eq. (30)

$$|v - \sum_{i=0}^{n_{win}-1} \tau_i(1 - Z_i)x|^2 \;\leq\; 2/L^2. \tag{53}$$

The vector $\tau_i(1 - Z_i)x$ is in $\mathcal{X}_i$. So, by Eq. (51),

$$\begin{aligned} |(1-P) \sum_{i=0}^{n_{win}-1} \tau_i(1 - Z_i)x| \;&\leq\; \text{const.} \times (\sqrt{F_0(l_b)}/l_b)\sqrt{\sum_{i=0}^{n_{win}-1} |\tau_i(1 - Z_i)x|^2} \\ &\leq\; \text{const.} \times (\sqrt{F_0(l_b)}/l_b)\sqrt{\sum_{i=0}^{n_{win}-1} |\tau_i x|^2} \\ &\leq\; \text{const.} \times (\sqrt{F_0(l_b)}/l_b). \end{aligned} \tag{54}$$

Combining Eqs. (53,54) with a triangle inequality verifies the first claim, given that $F(L)$ is chosen to grow slower than any power of L.

Proof of Second Claim— To prove the second claim **(2)**, consider any vector $v \in \mathcal{W}$. We have $v = \sum_i Ax_i$, with $x_i \in \mathcal{X}_i$ and $U \sum_i x_i = \sum_i x_i$, so $v = \sum_i AUx_i$. So, $|(1-P)Jv| = |\sum_i (1-P)JAUx_i|$. We have $|\sum_i (1-P)JAUx_i|^2 = \sum_{i,j}(JAUx_i, (1 - P)JAUx_j)$. Let R_k project onto the k-th block of the space $\mathcal{R}$. Note that $(1 - P)AUx_i = 0$, so $(1 - P)\omega(il_b)AUx_i = 0$, so

$$\begin{aligned} |\sum_i (1 - P)JAUx_i|^2 \;&=\; |(1 - P)\sum_i \Big(J - \omega(il_b)\Big)AUx_i)|^2 \\ &\leq\; |\sum_i \Big(J - \omega(il_b)\Big)AUx_i|^2 \\ &\leq\; \sum_{i,j} |(\Big(J - \omega(il_b)\Big)AUx_i, \Big(J - \omega(jl_b)\Big)AUx_j)|. \end{aligned} \tag{55}$$

Using the decay in lemma (4), the inner product above decays exponentially in $|i - j|$, so we can sum over i, j to find

$$(Jv, (1 - P)Jv) \leq \text{const.} \times (l_b\kappa)^2 \sum_i |x_i|^2. \tag{56}$$

By Eqs. (56,49), we have

$$\begin{aligned} |(1 - P)Jv|^2 \;&\leq\; \text{const.} \times (l_b\kappa)^2 \sum_i |x_i|^2 \\ &\leq\; \text{const.} \times (1/l_b^2)\Big(l_b\kappa\Big)^2 |v|^2 \\ &=\; \text{const.} \times \Big(l_b^2\kappa\Big)^2 |v|^2, \end{aligned} \tag{57}$$

verifying the second claim.

Proof of Third Claim— As we established before, using the Lieb-Robinson bound, for the given choice of $F(x)$ the norm of the projection of any vector $y \in \mathcal{X}_i$ onto $\mathcal{V}_L$ is bounded by $|y|$ times a function decaying faster than any negative power of L. Let $P_{\mathcal{V}_L}$ project onto $\mathcal{V}_L$. Using Eq. (49), we find that the projection of any vector $v \in \mathcal{W}$ onto $\mathcal{V}_L$ is bounded by (writing $v = \sum_{i=0}^{n_{win}-1} w_i$ with $w_i \in \mathcal{X}_i$)

$$\begin{aligned} |P_{\mathcal{V}_L} v|^2 \;&\leq\; n_{win} \sum_{i=0}^{n_{win}-1} |P_{\mathcal{V}_L} w_i|^2 \\ &\leq\; n_{win} \max_i(|P_{\mathcal{V}_L} w_i|^2/|w_i|^2)(1/l_b)^2|v|^2. \end{aligned} \tag{58}$$

Since $(|P_{\mathcal{V}_L} w_j|^2/|w_j|^2)$ is bounded by a function decaying faster than any negative power of L, this verifies the third claim.

This completes the proof of Lemma (2). After giving the error bounds in the next section, we explain some of the motivation behind the above construction, and comment on the easier case in which J is a tridiagonal matrix, rather than a block tridiagonal matrix.

V. ERROR BOUNDS

We finally give the error bounds to obtains theorems (1,2). To obtain (2), we pick

$$n_{cut} = \Delta^{-1/4}, \tag{59}$$

so that $L = \lfloor (2/n_{cut})/\Delta) - 1 \rfloor$ is of order $2/\Delta^{3/4}$. Then, from lemma (2) and Eq. (17), in the new basis the block-off-diagonal terms in H are bounded in operator norm by a constant times $\Delta^{1/4}$ times a function growing slower than any power of $1/\Delta$. By Eq. (18), the difference between B and B' is bounded in operator norm by a constant times $\Delta^{1/4}$. Therefore, theorem (2) follows. To obtain theorem (1), we pick

$$\Delta = \delta^{4/5} \tag{60}$$

in lemma (1).

We omit the detailed analysis, but it is possible to choose $E(x)$ to be a polylog as follows. We can pick $T(x)$ to decay like $\exp(-x^\eta)$, for any $\eta < 1$[29, 30]. Then we can pick $F(L)$ to equal $\log(L)^\theta$, for $\theta > 1/\eta$, so that $T(F(L)) \sim \exp(-(\log(L))^{\theta/\eta})$ decays faster than any power.

VI. TRIDIAGONAL MATRICES

In this section, we present tighter bounds for the case in which H is a tridiagonal matrix, rather than a block tridiagonal matrix.

Remark: The difficulty we face is that the $\mathcal{X}_i$ are not orthogonal to each other. If they were orthogonal, then many of the estimates would be easier. Consider the case in which J is a block diagonal matrix, so that $\mathcal{V}_1$ is one dimensional. Let $\rho(E)$ be a smoothed density of states at energy E: $\rho(E) = \mathrm{tr}(S^\dagger \mathcal{F}(E, 1/L, 1/L, J)^\dagger \mathcal{F}(E, 1/L, 1/L, J) S)$. Suppose $\rho(E)$ is such that it has a peak in the crossing points of Fig. 2a (the points where one function $\mathcal{F}$ is decreasing and the other is increasing and they cross). Then, with the overlapping windows as shown, we find that most of the smoothed density of states lies in the overlap between the windows, rather than in the windows themselves. The overlap between the vectors in different windows is large. In the case of a tridiagonal matrix, we can combine two of the windows as shown in Fig. 2b to reduce the overlap of the normalized vectors; this general idea will motivate the construction in this section.

We prove that

Lemma 5. *Let J be an L-by-L Hermitian tridiagonal matrix, with $\|J\| \leq 1$ acting on a space $\mathcal{B}$. Let v_j denote the vector with a 1 in the j-th entry and zeroes elsewhere. Then, there exists a space $\mathcal{W}$ which is a subspace of $\mathcal{B}$ with the following properties:*

- **(1)**: *The projection of v_1 onto the orthogonal complement of $\mathcal{W}$ has norm bounded by ϵ_3 where ϵ_3 is equal to a constant times $1/L$.*
- **(2)**: *For any normalized vector $w \in W$, the projection of Jw onto the orthogonal complement of $\mathcal{W}$ has norm bounded by ϵ_4, where ϵ_4 is equal to $1/L$ times a function growing slower than any power of L.*
- **(3)**: *The projection of v_L onto $\mathcal{W}$ has norm bounded by ϵ_5, where ϵ_5 is a function decaying faster than any power of L.*

This lemma implies theorem (3): we construct A', B' as before, following steps **(3)** to construct the new basis, but because of the tighter bounds in lemma (5) we can choose $n_{cut} = \Delta^{-1/2}$ when constructing the new basis. Now, in step **(4)**, we find that A', B' are diagonal matrices, rather than just block diagonal matrices.

For each $i = 0, 1, ..., n_{win} - 1$, define

$$\omega(i) = -1 + i\kappa, \tag{61}$$

as before. Define

$$\rho_i = \mathrm{tr}\Big(S^\dagger \mathcal{F}(\omega(i),0,\kappa,J)^\dagger \mathcal{F}(\omega(i),0,\kappa,J) S\Big) \quad (62)$$
$$= \Big|\mathcal{F}(\omega(i),0,\kappa,J) v_1\Big|^2.$$

Set

$$\lambda_{min} = 1/(n_{win} L^2), \quad (63)$$

as before with

$$n_{win} = L/F(L) \quad (64)$$

as before. To prove Lemma (5), we use the following algorithm. There are n_{win} windows, labeled $0,...,n_{win}-1$. We label various windows as either "unmarked" or "marked"; windows which are marked get marked by an integer label.

1: Set $i=0$. Initialize a real variable x to 0. Initialize an integer counter a to 1. Initialize all windows to unmarked.

2: Set x to 0. **If** $\rho_i < \lambda_{min}$,

then

2a: Increment i by one.

2b: If $i \geq n_{win}$, terminate. Otherwise, go to step **2**.

endif

3: Mark window i with label a.

4: Set x to $x+\rho_i$. **If** $x < 9\rho_i$,

then

4a: Increment i by one.

4b: If $i \geq n_{win}$, terminate. Otherwise, go to step **3**.

endif

5: Increment a by one. Increment i by one. If $i \geq n_{win}$, terminate. Otherwise, goto step **2**.

After running this algorithm, there will be a sequences of marked windows all marked with the same integer label a. There may be one or more unmarked windows separating the sequences of marked windows. In step 2, we scan along to find an i with $\rho_i \geq \lambda_{min}$, and then in step 4 we mark a sequence of windows. We claim that the length of a sequence of marked windows is at most $1+\lceil \log_{10/9}(2/\lambda_{min})\rceil$. This bound on the length of a sequence of marked windows holds because at the start of a sequence x is at least λ_{min}, x grows exponentially along the sequence (otherwise in step **4** we find that $\rho_{i+1} > (1/9)x$ for some i), and x can be at most 2 since $\sum_{i=0}^{n_{win}-1} \rho_i \leq 2$.

Let the total number of sequences be n_{seq}. Note that $n_{seq} \leq n_{win}$.

For each sequence of windows marked with a given integer a, from window i to j, construct the vector y_a given by

$$y_a = \sum_{k=i}^{j} \mathcal{F}(\omega(k),0,\kappa,J) v_1. \quad (65)$$
$$= \mathcal{F}((\omega(i)+\omega(j))/2, (\omega(j)-\omega(i))/2, \kappa, J) v_1.$$

The inner product (y_a, y_{a+1}) is equal to $(\mathcal{F}(\omega(j),0,\kappa,J)v_1, y_{a+1})$. By Cauchy-Schwarz, this is bounded by $|(\mathcal{F}(\omega(j),0,\kappa,J)v_1||y_{a+1}|$. To estimate $|(\mathcal{F}(\omega(j),0,\kappa,J)v_1|$, we use $|(\mathcal{F}(\omega(j),0,\kappa,J)v_1|^2 = \rho_j \leq \sum_{k=i}^{j} \rho_k/9 \leq |y_a|^2/9 = \sum_{k=i}^{j}\sum_{k'=i}^{j} (\mathcal{F}(\omega(k),0,\kappa,J)v_1, \mathcal{F}(\omega(k'),0,\kappa,J)v_1)/9$, where the first inequality is by construction and the second inequality follows from the fact that $(\mathcal{F}(\omega(k),0,\kappa,J)v_1, \mathcal{F}(\omega(k'),0,\kappa,J)v_1) \geq 0$. Therefore, $(y_a, y_{a+1}) \leq (|y_a|/\sqrt{9})|y_{a+1}|$, so

$$(y_a, y_{a+1}) \leq (1/3)|y_a||y_{a+1}|. \quad (66)$$

We define $\mathcal{W}$ to be the space spanned by all such vectors y_a, and we define P to project onto $\mathcal{W}$. Consider any vector $v \in \mathcal{W}$, with

$$v = \sum_{a=1}^{n_{seq}} v_a, \tag{67}$$

with v_a parallel to y_a. By Eq. (66)

$$|v|^2 \geq \frac{1}{3} \sum_{a=1}^{n_{seq}} |v_a|^2. \tag{68}$$

Remark: The function $\mathcal{F}((\omega(i)+\omega(j))/2, (\omega(j)-\omega(i))/2, \kappa, \omega)$ is equal to unity for $\omega(i) \leq \omega \leq \omega(j)$.

We now prove the Lemma (5) as follows: to prove the first claim, note that by construction,

$$\begin{aligned} |Pv_1 - v_1|^2 &\leq |\sum_{a=1}^{n_{seq}} y_a - v_1|^2 \\ &\leq 2 n_{win} \lambda_{min} \\ &\leq 2/L^2. \end{aligned} \tag{69}$$

The second line of the above equation follows because the difference $\sum_a y_a - v_1$ is equal to $-\sum_{i \text{unmarked}} \mathcal{F}(\omega(i), 0, \kappa, J) v_1$, where the sum ranges over i such that the corresponding window is unmarked.

To prove the second claim, consider the a-th sequence of marked windows, from window i to window j. Let $\omega_a = (\omega^-(i) + \omega^+(j))/2$. Then,

$$|(J - \omega_a) y_a| \leq \Big(\frac{2 + \lceil \log_{10/9}(2/\lambda_{min}) \rceil}{n_{win}} \Big) |y_a| \tag{70}$$

which is bounded by $1/L$ times a function growing slower than any power of L. Therefore,

$$|(1-P) J y_a| \leq \Big(\frac{2 + \lceil \log_{10/9}(2/\lambda_{min}) \rceil}{n_{win}} \Big) |y_a| \tag{71}$$

Using the bound Eq. (68), for any vector $v \in \mathcal{W}$,

$$|(1-P) J v| \leq 2\sqrt{3} \Big(\frac{2 + \lceil \log_{10/9}(2/\lambda_{min}) \rceil}{n_{win}} \Big) |v|, \tag{72}$$

which is bounded by $1/L$ times a function growing slower than any power of L, verifying the second claim.

The proof of the third claim is identical to the previous case.

VII. QUANTUM MEASUREMENT

A. Construction and Results

The constructions above can be applied to operators which arise in various physical quantum systems. For example, consider a quantum spin for a large spin S. Then, the operators S_x/S and S_y/S have operator norm 1 and have a commutator that is of order $1/S$. Thus, we can find a basis in which both operators are almost diagonal. While it is well known that one can use a POVM (positive operator-valued measure) to approximately measure S_x and S_y at the same time, the existence of the given basis implies that one can approximately measure S_x and S_y simultaneously with a single *projective* measurement. Interestingly, while the operator S_z^2 is also almost diagonal in this basis (since it equals $S(S+1) - S_x^2 - S_y^2$), it is not possible to find a basis in which S_x, S_y, and S_z are all almost diagonal (this obstruction is similar to that in [6]). Therefore, to approximately measure S_x, S_y, and S_z simultaneously will require a POVM, rather than a projective measurement.

For completeness, we now briefly show how to construct a POVM to approximately measure several almost commuting operators simultaneously. Consider any number N of Hermitian matrices, labeled $A_1, ..., A_N$, with $\|[A_i, A_j]\| \leq \delta$

for all i, j and with $\|A_i\| \leq 1$ for all i. We now construct a POVM to approximately measure all N operators simultaneously. The physical idea is very simple: we first do a “soft” measurement of A_N, then A_{N-1}, and so on, until all operators are measured.

Let n_{win} be some integer given by

$$n_{win} = \lceil \delta^{-1/2}(N-1)^{-1/2} \rceil \tag{73}$$

(n_{win} will typically be much larger than unity). For $i = 1, ..., N$ and $n = 0, ..., n_{win} - 1$, define

$$\omega(i) = -1 + 2i/(n_{win} - 1) \ = \ -1 + i\kappa, \tag{74}$$

where $\kappa = 2/(n_{win} - 1)$ as before, and define

$$M(i, n) = \sqrt{\mathcal{F}(\omega(n), 0, \kappa, A_i)}. \tag{75}$$

The definition of $\mathcal{F}$ is given at the start of section IV; we will see later that in this section that we do not actually need $\mathcal{F}$ to be infinitely differentiable as it is defined there, but we have only weaker requirements on $\mathcal{F}$. Define

$$O(n_1, n_2, ..., n_N) = \Big(M(1, n_1)^\dagger M(2, n_2)^\dagger ... M(N, n_N)^\dagger\Big)\Big(M(N, n_N) ... M(2, n_2) M(1, n_1)\Big). \tag{76}$$

Then,

$$\sum_{n_1, n_2, ... = 0}^{n_{win}-1} O(n_1, n_2, ..., n_N) = \mathbb{1}, \tag{77}$$

and all of the operators $O(n_1, n_2, ..., n_N)$ are positive semidefinite by construction. Therefore, the operators $O(n_1, n_2, ..., n_N)$ form a POVM. Note that $M(i, n_i) = M(i, n_i)^\dagger$, but we continue to write daggers on the operators for clarity.

We claim that this POVM approximately measures all operators simultaneously. That is, we will show that for any density matrix ρ, if the outcome of the measurement is $n_1, n_2, ..., n_N$, then if we perform a subsequent measurement of any operator A_i, the outcome will be close to $\omega(n_i)$ with high probability. We show this by computing the expectation value $(A_i - \omega(n_i))^2$ averaged over all measurement outcomes. For any density matrix ρ, for any i, the average over all outcomes of $(A_i - \omega(n_i))^2$ is equal to

$$\sum_{n_1, n_2, ... = 0}^{n_{win}-1} \mathrm{tr}\Big((A_i - \omega(n_i))^2 M(1, n_1) M(2, n_2) ... \rho ... M(2, n_2)^\dagger M(1, n_1)^\dagger\Big) \tag{78}$$

The main result in this section is that

$$\sum_{n_1, n_2, ... = 0}^{n_{win}-1} \mathrm{tr}\Big((A_i - \omega(n_i))^2 M(1, n_1) M(2, n_2) ... \rho ... M(2, n_2)^\dagger M(1, n_1)^\dagger\Big) \leq \text{const.} \times (N-1)\delta. \tag{79}$$

We show this in the next subsection.

B. Bounds

Note that $\| \sum_{n_i} M(i, n_i) \| \leq \sqrt{2}$. To bound Eq. (78), we need three results, Eqs. (80,81,82) below. First,

$$\begin{aligned} \sum_{n_i=0}^{n_{win}-1} \|(A_i - \omega(n_i)) M(i, n_i)\| \ &\leq \ \text{const.} \times \kappa \\ &\leq \ \text{const.} \times 1/n_{win}. \end{aligned} \tag{80}$$

Second, we need

$$\| \sum_{n_j=0}^{n_{win}-1} [M(j, n_j), (A_i - \omega(n_i))] O M(j, n_j)^\dagger \| \leq \text{const.} \times (\delta/\kappa) \|O\| \tag{81}$$

for any operator O.

Third, we need

$$\|\sum_{n_j=0}^{n_{win}-1}[M(j,n_j),(A_i-\omega(n_i))]O[M(j,n_j)^\dagger,(A_i-\omega(n_i))]\| \leq \text{const.}\times(\delta/\kappa)^2\|O\| \tag{82}$$

for any operator O.

Eq. (80) follows immediately from the support of $\mathcal{F}$. To show Eq. (81), define

$$A^0=\kappa\int \mathrm{d}t \exp(iA_jt)(A_i-\omega(n_i))\exp(-iA_jt)f(\kappa t), \tag{83}$$

where the function $f(t)$ is defined to have the Fourier transform as in Eq. (5). Then, $\|A^0-(A_i-\omega(n_i))\|\leq \text{const.}\times\delta/\kappa$ as in lemma (1). Also, if v_1,v_2 are eigenvectors of A_j with corresponding eigenvalues x_1,x_2 with $|x_1-x_2|\geq\kappa$, then $(v_1,A^0v_2)=0$, which implies that

$$\|\sum_{n_j=0}^{n_{win}-1}[M(j,n_j),A^0]OM(j,n_j)^\dagger\|\leq 2\max_{n_j}(\|[M(j,n_j),A^0]OM(j,n_j)^\dagger\|). \tag{84}$$

Eq. (84) is the reason for introducing the operator A^0. We can bound the commutator $[M(j,n_j),A^0]$ as follows. Note that $\|[A_j,A^0]\|\leq\delta$. Write

$$M(j,n_j)=\int \mathrm{d}t \exp(iA_jt)\widetilde{\sqrt{\mathcal{F}(\omega(n_j),0,\kappa,t)}}, \tag{85}$$

where $\widetilde{\sqrt{\mathcal{F}(.,.,.,t)}}$ denotes the Fourier transform of the square-root of $\mathcal{F}$. Then since $\|[\exp(iA_jt),A^0]\|\leq\text{const.}\times|t|\delta$, we can use a triangle inequality to show that

$$\|[M(j,n_j),A^0]\|\leq\int \mathrm{d}t\,\widetilde{\sqrt{\mathcal{F}(\omega(n_j),0,\kappa,t)}}|t|\delta. \tag{86}$$

Then, since $\sqrt{\mathcal{F}}(.,.,.,\omega)$ is infinitely differentiable, the Fourier transform decays faster than any power of t and the integral over t converges, so we have $\|[M(j,n_j),A^0]\|\leq\text{const.}\times\delta/\kappa$. Using Eq. (84) gives Eq. (81). Eq. (82) is derived similarly.

Using Eqs. (80,81,82), we can bound the sum in Eq. (78) by writing $(A_i-\omega(n_i))^2=(A_i-\omega(n_i))(A_i-\omega(n_i))$, and commuting one of the terms $(A_i-\omega(n_i))$ to the right through $M(j,n_j)$ for $j<i$ until it hits the $M(i,n_i)$ and commuting the other term $(A_i-\omega(n_i))$ to the left through $M(j,n_j)^\dagger$ for $j<i$ until it hits $M(i,n_i)^\dagger$. Therefore,

$$\begin{aligned}
&\sum_{n_1,n_2,\ldots=0}^{n_{win}-1}\mathrm{tr}((A_i-\omega(n_i))^2M(1,n_1)M(2,n_2)...\rho...M(2,n_2)^\dagger M(1,n_1))^\dagger \qquad (87)\\
\leq\;& \text{const.}\times\Big((i-1)^2\delta^2n_{win}^2+(i-1)\delta n_{win}/n_{win}+1/n_{win}^2\Big)\\
\leq\;& \text{const.}\times\Big((N-1)^2\delta^2n_{win}^2+(N-1)\delta+1/n_{win}^2\Big).
\end{aligned}$$

The first term on the right-hand side of Eq. (87) arises from two non-vanishing commutators (if the non-vanishing commutators are with $M(j,n_j)$ and $M(k,n_k)^\dagger$ for $j\neq k$ then we use Eq. (81) twice, but if $j=k$ we use Eq. (82) once). The second term arises from one non-vanishing commutator and one use of Eq. (80), and the last term arises from using Eq. (80) twice. Choosing

$$n_{win}=\lceil\delta^{-1/2}(N-1)^{-1/2}\rceil, \tag{88}$$

we find that we measure all operators to within a mean-square error of order $(N-1)\delta$, as claimed.

Note that we did not actually require that $\mathcal{F}(.,.,.,\omega)$ be infinitely differentiable in this section. We only required that the Fourier transform $\widetilde{\sqrt{\mathcal{F}(.,.,.,t)}}$ decay sufficiently rapidly in t that the integral (86) converges. The other properties of $\mathcal{F}$ we used are that $\sum_n\mathcal{F}(\omega(n),0,\kappa,\omega)=1$ for $-1\leq\omega\leq1$ and that $\mathcal{F}(\omega(n),0,\kappa,\omega)$ vanish for $|\omega-\omega(n)|\geq\kappa$.

VIII. DISCUSSION

The main result is an explicit construction of a pair of exactly commuting matrices which are close to a pair of almost commuting matrices. The construction of the matrix is explicit and can be handled easily on a computer for modest N. We have in fact implemented the construction in Lemma (5) for the uniform chain. In practical applications, we expect that, for many tridiagonal matrices, the lack of orthogonality of the $\mathcal{X}_i$ will not cause a problem, and choosing $\mathcal{W}$ to be the space spanned by the $\mathcal{X}_i$ will lead to satisfactory results, without having to follow the more complicated procedure above. If, for some particular J, the lack of orthogonality of the $\mathcal{X}_i$ does cause a problem, an alternative procedure that might be more useful in practice than the deterministic procedure above is to add small, randomly chosen matrices to each diagonal block of J. This may smooth out the spectrum of J and then allow one to choose $\mathcal{W}$ to be the space spanned by the $\mathcal{X}_i$.

We gave above applications to quantum measurement. Another application of this result is to construct Wannier functions for any two dimensional quantum system for a spectral gap. In [31], it was pointed out that given a two dimensional quantum system with a gap between bands, one could define an operator G which projected onto the bands below the gap. Then, define the operator X and Y to measure X and Y position of particles, and define GXG and GYG as projections of X and Y into the lowest band. Let $\|X\|, \|Y\| = L$, where L is the linear size of the system. Since the operator G was constructed in [31] as a short-range operator, the commutator $\|[GXG, GYG]\|$ is small compared to L^2, and thus we can use the results here to construct a basis of Wannier functions which is localized in both the x- and y-directions.

Acknowledgments— I thank Y. Ogata for pointing out a gap in the proof on the previous draft in the proof of the "second property" (now the first property) and for many very useful discussions; while filling this gap in, I found the modification of the construction here which in fact leads to tighter bounds. I thank G. Bouch for useful comments. I thank T. J. Osborne and J. Yard for useful conversations, and I thank T. J. Osborne for many comments on a draft of this paper. I thank D. Poulin for explaining Jordan's lemma. This work was supported by U. S. DOE Contract No. DE-AC52-06NA25396.

[1] P. R. Rosenthal: Are almost commuting matrices near commuting pairs?. AMS Monthly **76**, 925 (1969).
[2] P. R. Halmos: Some unsolved problems of unknown depth about operators on Hilbert space. Proc. Roy. Soc. Edinburgh A **76**, 67 (1976).
[3] H. Lin: Almost commuting self-adjoint matrices and applications. Fields. Inst. Commun. **13**, 193 (1995).
[4] P. Friis and M. Rordam: Almost commuting self-adjoint matrices—a short proof of Huaxin Lin's theorem. J. Reine Angew. Math. **479**, 121 (1996).
[5] D. Voiculescu: Asymptotically commuting finite rank unitaries without commuting approximants. Acta Sci. Math. **451**, 429 (1983).
[6] D. Voiculescu: Remarks on the singular extension in the C*-algebra of the Heisenberg group. J. Op. Thy. **5**, 147 (1981).
[7] K. R. Davidson: Almost commuting Hermitian matrices. Math. Scand. **56**, 222 (1985).
[8] E. H. Lieb and D. W. Robinson: The finite group velocity of quantum spin systems. Commun. Math. Phys. **28**, 251 (1972).
[9] M. B. Hastings: Lieb-Schultz-Mattis in higher dimensions. Phys. Rev. B **69** 104431 (2004).
[10] M. B. Hastings and T. Koma: Spectral gap and exponential decay of correlations. Commun. Math. Phys. **265**, 781 (2006).
[11] B. Nachtergaele and R. Sims: Lieb-Robinson bounds and the exponential clustering theorem. Commun. Math. Phys. **265**, 119 (2006).
[12] S. Demko, W. F. Moss, P. W. Smith: Decay rates for inverses of band matrices. Math. Comp., **43**, 491 (1984).
[13] M. Benzi and G. H. Golub: Bounds for the entries of matrix functions with applications to preconditions. BIT, **39**, 417 (1999).
[14] M. B. Hastings: Locality in Quantum and Markov Dynamics on Lattices and Networks. Phys. Rev. Lett. **93**, 140402 (2004).
[15] M. Cramer, C. M. Dawson, J. Eisert, and T. J. Osborne: Exact relaxation in a class of nonequilibrium lattice systems. Phys. Rev. Lett. **100**, 030602 (2008).
[16] J. Eisert and T. J. Osborne: General entanglement scaling laws from time evolution. Phys. Rev. Lett. **97**, 150404 (2006).
[17] S. Bravyi, M. B. Hastings, and F. Verstraete: Lieb-Robinson Bounds and the Generation of Correlations and Topological Quantum Order. Phys. Rev. Lett. **97**, 050401 (2006).
[18] T. J. Osborne: A Renormalization-Group Algorithm for Eigenvalue Density Functions of Interacting Quantum Systems. arXiv:cond-mat/0605194.
[19] T. J. Osborne: Efficient Approximation of the Dynamics of One-Dimensional Quantum Spin Systems. Phys. Rev. Lett. **97**, 157202 (2006).
[20] M. B. Hastings: Quantum Belief Propagation. Phys. Rev. B Rapids **76**, 201102 (2007).
[21] T. J. Osborne: Efficient approximation of the dynamics of one-dimensional quantum spin systems. Phys. Rev. A **75**, 042306 (2007).

[22] M. B. Hastings: Observations Outside the Light-Cone: Algorithms for Non-Equilibrium and Thermal States. Phys. Rev. B **77**, 144302 (2008).
[23] M. B. Hastings: An Area Law for One Dimensional Quantum Systems. J. Stat. Mech., P08024 (2007).
[24] B. Nachtergaele, H. Raz, B. Schlein, and R. Sims: Lieb-Robinson Bounds for Harmonic and Anharmonic Lattice Systems. arXiv:0712.3820.
[25] M. B. Hastings: Quasi-Adiabatic Continuation in Gapped Spin and Fermion Systems: Goldstone's Theorem and Flux Periodicity. J. Stat. Mech., P05010 (2007).
[26] R. Carmona, A. Klein, and F. Martinelli: Anderson localization for Bernoulli and other singular potentials. Commun. Math. Phys. **108**, 41 (1987).
[27] H. von Dreifus and A. Klein: A new proof of localization in the Anderson tight binding model. Commun. Math. Phys. **124**, 285 (1989). C. Shubin, R. Vakilian, and T. Wolff, Some harmonic analysis questions suggested by Anderson-Bernoulli models. Geom. Funct. Anal. **8**, 932 (1998).
[28] Jordan's lemma states that given any two projectors, M and N, we can find a basis in which both projectors are block diagonal with blocks of size 1 or 2. Equivalently (this is the formulation we use), if M projects onto a subspace $\mathcal{M}$, then we can find an orthonormal basis of vectors m_i for $\mathcal{M}$ with the property that $(m_i, Nm_j) = 0$ for $i \neq j$. The original reference is C. Jordan: Essai sur la géométrie à n dimensions. Bulletin de la S. M. F. **3**, 103 (1875).
[29] R. Gervais, Q. I. Rahman, and G. Schmeisser: A bandlimited function simulating a duration-limited one, *Anniversary volume on approximation theory and functional analysis*, Schiftenreihe Numer. Math., 355 (1984).
[30] T. Strohmer and J. Tanner: Implementations of Shannon's sampling theorem, a time-frequency approach. Sampling Thy. in Signal and Image Processing, **4**, 1 (2005); Fritz John, *Partial Differential Equations*, Springer-Verlag, New York, 1991.
[31] M. B. Hastings: Topology and Phases in Fermionic Systems. J. Stat. Mech., L01001 (2008).